TECHNICAL NOTE

Ranajit Chakraborty,[1] Ph.D. and David N. Stivers,[1] Ph.D.

Paternity Exclusion by DNA Markers: Effects of Paternal Mutations

REFERENCE: Chakraborty, R. and Stivers, D. N., "**Paternity Exclusion by DNA Markers: Effects of Paternal Mutations,**" *Journal of Forensic Sciences*, JFSCA, Vol 41, No. 4, July 1996, pp. 671–677.

ABSTRACT: In parentage testing when one parent is excluded, the distribution of the number of loci showing exclusion due to mutations of the transmitting alleles is derived, and it is contrasted with the expected distribution when the exclusion is caused by non-paternity. This theory is applied to allele frequency data on short tandem repeat loci scored by PCR analysis, and VNTR data scored by Southern blot RFLP analysis that are commonly used in paternity analysis. For such hypervariable loci, wrongly accused males should generally be excluded based two or more loci, while a true father is unlikely to be excluded based on multiple loci due to mutations of paternal alleles. Thus, when these DNA markers are used for parentage analysis, the decision to infer non-paternity based on exclusions at two or more loci has a statistical support. Our approach places a reduced weight on the combined exclusion probability. Even with this reduced power of exclusion, the probability of exclusion based on combined tests on STR and VNTR loci is sufficiently large to resolve most paternity dispute cases in general populations.

KEYWORDS: forensic science, paternity testing, STR, VNTR, germline mutation

Introduction

It is now well known that in parentage testing, DNA markers are more efficient than the classical markers, both in terms of exclusionary power for wrongly accused men and inclusion probability for identification of the true biological father (1–5).

As a consequence, the paternity testing laboratories in the US and elsewhere are in the process of supplementing and/or replacing the traditional serologic markers with DNA markers. In a 1990 survey by the American Association of Blood Banks (AABB) conducted on 34 paternity testing laboratories of the US, nearly 10% of the paternity dispute cases were reported to have been resolved by DNA testing only (1). In a similar survey also conducted by the AABB, 50 laboratories reported to have used DNA testing in nearly 40% of the case-work analyses (Dr. J. W. Morris, personal communication).

The high efficiency of DNA markers is not the only reason for using these polymorphic markers as a replacement of the standard serological markers for resolving paternity disputes. As discussed by Alford et al. (5) and Pena and Chakraborty (2), DNA testing can also be used to resolve paternity from hairs, paraffin-included blocks of tissue, and even exhumed materials in cases of deceased parents; these are situations in which the classical serological markers are difficult, if not impossible, to use. The high efficiency of DNA markers, of course, has its attendant caveats. This efficiency arises from their comparatively higher number of segregating alleles and consequently higher per-locus heterozygosity. These biological features of DNA markers are caused by a higher rate of mutation at DNA loci in comparison with the traditional markers. In particular, the DNA markers that are most efficient for personal identification purposes are those whose polymorphism is due to variation of the copy numbers of short tandemly repeated nucleotide sequences in specific regions of the genome (called the Variable Number of Tandem Repeat, VNTR, loci). Jeffreys et al. (6) reported empirical data on the rate of mutations at such loci, indicating that the mutation rate at such loci increases with per-locus variability. As a consequence, mutations may become a major factor in deciding whether or not sporadic discordances of genotypes between a child and its putative father could truly be due to mutations, and not due to non-paternity.

A 1991 recommendation by the AABB suggests that when hypervariable single-locus probes are used for paternity determination, exclusion of paternity should be based on exclusionary events at 2 or more loci. While this suggestion is intuitively reasonable, it should accompany a rigorous statistical validation. The purpose of this study is to provide such a validation, using a theory proposed earlier (7,8). Specifically, we show that under certain justifiable assumptions, the distribution of the number of loci that would exclude a randomly accused man can be computed for every mother-child pair, which in turn can be contrasted with the expected distribution when exclusions occur only by mutations of paternal alleles. Using population data on eleven tandem repeat (STR) loci (HUMRENA4[ACAG]$_n$, HUMFESPFS[AAAT]$_n$, HUMFABP[AAT]$_n$, HUMCD4[AAAAG]$_n$, HUMCSF1PO[AGAT]$_n$, HUMTHO1[AATG]$_n$, HUMPLA2A1[AAT]$_n$, HUMF13A01[AAAAG]$_n$, HUMCYAR04[AAAT]$_n$, HUMTRINRAA[TAA]$_n$ and HUMLIPOL[AAAT]$_n$), we show a clear dichotomy of the distribution of the number of loci at which exclusions are found for excluded random men, and the one expected from mutations for the STR loci. In contrast, with the possibility of mutations, a single-locus exclusion could become frequent enough for the battery of VNTR probes used in RFLP analysis for paternity testing. Thus, the inference of true exclusion, based on excluding a parent when at least two exclusionary events

[1]Human Genetics Center, School of Public Health, The University of Texas-Houston Health Science Center.

Received for publication 8 Sept. 1995; revised manuscript received 21 Dec. 1995; accepted for publication 2 Jan. 1996.

Copyright © 1996 by ASTM International

(as suggested by the AABB), appears reasonable for the battery of single-locus VNTR probes. Finally, the implications of these findings in selecting a battery of DNA typing loci for paternity analysis are addressed to revisit the question of relative utility of multiple numbers of single locus probes versus multilocus DNA fingerprint methods in parentage analysis (9,10).

Theory

Analysis without invoking mutations

For L unlinked loci, let $\overline{P}_E(l)$ denote the locus-specific probability of exclusion for the l-th locus ($l = 1, 2, \ldots L$). In other words, $\overline{P}_E(l)$ is the chance that a random alleged father would be excluded based on DNA typing at the l-th locus. It is well documented that this probability can be computed at a locus-level averaged over all possible mother-child pair (11–13), or for each possible combinations of mother-child genotype pairs (14). For example, for a multi-allelic codominant locus, the average exclusion probability is given by

$$\overline{P}_E(l) = 1 - 2a_2 + a_3 + 3(a_2a_3 - a_5) - 2(a_2^2 - a_4) \quad (1)$$

where a_r is the sum of r-th power of all allele frequencies at a locus (12,13). In contrast, for a specific mother-child pair, the exclusion probability for the same locus takes the form

$$P_E(l) = \begin{cases} (1 - p_i)^2 \\ (1 - p_i - p_j)^2 \end{cases} \quad (2)$$

where the first event refers to the case where the child can receive a single specific allele (say, A_i) from the biological father, and the second event occurs when both the mother and child are heterozygous for the same pairs of alleles (say A_i and A_j). Thus, for any given i and j, the first event takes place with probability $p_i(1 - p_i + p_i^2)$ and the second with probability $p_ip_j(p_i + p_j)$, for all $i = 1, 2, \ldots k$ and $j > i$ (13,14). The implicit assumptions in deriving equations (1) and (2) are that the population is in approximate Hardy-Weinberg equilibrium at the locus, and that the allele frequencies in the male and female gene pool of the population are equal. When the mother and father are presumably from different populations, the respective expressions are given in (13). It is important to note that most of the discussions of utility of genetic markers in the context of parentage analysis relate to the average probability of exclusion offered by the locus (equation 1). However, in reality the observed exclusion rates should use (2), since these probabilities may be substantially different from the average for the locus as a whole.

When L such loci are used in paternity analysis, the combined probability of exclusion is given by

$$P_E(C) = 1 - \prod_{l=1}^{L} (1 - P_E(l)) \quad (3)$$

which may again be computed by replacing the locus-specific values for an average mother-child pair (equation 1), or for the specific mother-child genotype pair observed at the loci (equation 2). The additional assumption in deriving equation (3) is that the alleles at different loci segregate independently of each other (i.e., the loci are under gametic phase equilibrium (15,16)). Also implicit in equation (3) is the notion that a random male is excluded when at least one locus shows exclusion of paternity.

Regardless of the specific formula used, when the $P_E(l)$ values are available for L unlinked loci ($l = 1, 2, \ldots L$), the logic of derivation for the distribution of the number of loci with respect to which exclusions would be found for a random alleged male was first given by Chakraborty and Schull (7), and used in other contexts in Chakraborty (8). A mathematically equivalent, and computationally simpler, version of their logic may be described as follows. Let Q_r^m be the probability that a randomly tested male shows exclusions based on exactly r of the first m loci tested ($r = 0, 1, 2, \ldots m$; $m = 1, 2, \ldots L$). The probability Q_r^L for $r = 0, 1, 2, \ldots L$ may be computed by using the recurrence relationship

$$Q_r^{m+1} = Q_r^m[1 - P_E(m + 1)] + Q_{r-1}^m P_E(m + 1) \quad (4)$$

for $m = 1, 2, \ldots L - 1$; and $r = 1, 2, \ldots m$ with the boundary conditions that for any $m = 1, 2, \ldots L$, we have

$$Q_0^m = \prod_{l=1}^{m} [1 - P_E(l)] \quad (5)$$

and

$$Q_m^m = \prod_{l=1}^{m} P_E(l) \quad (6)$$

Note that the values of $Q_0^L, Q_1^L, Q_2^L, \ldots Q_L^L$, computed from recursive use of equation (4) with boundary conditions (5) and (6) also give other important values that can be reported for paternity test results. For example, the combined power of exclusion based on L loci, when paternity exclusion is based on one or more locus-specific exclusions is $P_E(C) = 1 - Q_0(L)$, while combined power of exclusion of L loci when two or more locus-specific exclusions would be considered as the criterion of exclusion is $P_E^*(C) = 1 - Q_0^L - Q_1^L$. Now, note that

$$Q_1^L = \sum_{l=1}^{L} P_E(l) \prod_{l' \neq l} [1 - P_E(l')]$$
$$= Q_0^L \sum_{l=1}^{L} \frac{P_E(l)}{1 - P_E(l)}$$

Thus, the combined exclusionary power for L loci based on two or more exclusions, $P_E^*(C)$, becomes

$$P_E^*(C) = 1 - Q_0^L\left[1 + \sum_{l=1}^{L} \left(\frac{P_E(l)}{1 - P_E(l)}\right)\right] \quad (7)$$

Furthermore, the mean and variance of the number of loci showing exclusions are given by

$$\sum_{l=1}^{L} P_E(l) \quad (8)$$

and

$$\sum_{l=1}^{L} P_E(l)[1 - P_E(l)] \quad (9)$$

respectively, which may be computed for any specific choice of sets of locus-specific exclusion probabilities (equation 1 or 2). The

inclusion of X-linked or Y-linked loci also does not affect such computations (17), although for Y-linked loci $P_E(l)$ will be zero for a female child. For a X-linked locus, likewise, it would be zero for a male child. In equations (7)–(9), we further assume that the loci are mutually independent to each other (i.e., no linkage disequilibrium exists between loci).

Analysis Invoking Mutations

Thus far we have assumed that any discordance between parental and child's genotypes is due to non-paternity, and that the mother is always correctly identified. When hypervariable DNA markers are used, these assumptions are not necessarily valid, since for tandem repeat loci, mutation rates are occasionally high enough to produce genotypic discordances between parents and offspring even when the family relationship is correct. Rothman et al. (18) addressed the question of assigning the probability of paternity in the presence of mutations. However, the emphasis of their approach was to provide estimates of mutation rates in the presence of non-paternity in the population, whereas we are interested in the question of determining how often exclusion at one or more loci are to be expected in the presence of mutations. In this section, we assume that the mother and alleged father are the true biological parents of the tested child. Furthermore, we assume that the maternal allele is faithfully transmitted to the child, so that no mutation has occurred during the maternal meiosis. Under this scenario, for a specific mother-child pair, the probability that a true father will be excluded due to mutation of the paternal allele(s) is given by

$$P_E(\mu) = \begin{cases} \dfrac{\mu_i(1 - p_i)^2}{p_i} \\[2ex] \dfrac{(\mu_i + \mu_j)(1 - p_i - p_j)^2}{p_i + p_j} \end{cases} \qquad (10)$$

where, as in equation (2), the first probability applies when the child has received the i-th allele from the biological father, and the second probability applies when the mother and the child are heterozygous for the same pair of alleles (say, A_i and A_j). The probability of mutation to the i-th allele is μ_i. In general, μ_i can be assumed to be μ/k, where μ is the overall mutation rate for this locus and k is the number of allele at the locus.

Averaging over all mother-child genotypes pairs, the probability that the true biological father will be excluded due to mutation of paternal alleles is given by

$$\overline{P}_E(\mu) = \mu - \frac{\mu}{k}[2 + a_2 - 3a_3 + 3a_4 - 2a_2^2] \qquad (11)$$

The distribution of the number of loci showing exclusions where they are caused by mutations of paternal alleles when the alleged father is truly the biological father is also derived by the same theory, where the $P_E(l)$ values are replaced by equations (10) or (11). Equations (4)–(9) hold, giving a contrast of the distributions of number of loci showing exclusions by non-paternity versus by mutations of paternal alleles.

Application to Paternity Testing Cases Typed by Minisatellite VNTR and STR Loci

STR Loci

Edwards et al. (19) and Hammond et al. (20) characterized the allele frequency distributions at 13 short tandem repeat (STR) loci in four major human population groups of the USA (Caucasians, Afro-Americans, Hispanics, and Asians). These authors also estimated the locus-specific mutation rates at these loci. Of these two are X-linked loci (HUMHPRTB[AGAT]$_n$, HUMARA[AGC]$_n$), and hence, they are not generally used in parentage testing. While the mutation rates estimates given by these authors are indirect, good agreement of these estimates are found for loci where direct estimates are available (21). Table 1 presents the locus-specific exclusion probabilities at the 11 autosomal STR loci using the Caucasian allele frequencies from these surveys. Two points are worth noting from these computations. First, for each locus, the locus-specific probability of excluding a random male for a specific mother-child pair can vary substantially depending on the genotypes observed in the mother-child pair. This is evident from the differences of minimum and maximum values of exclusion probabilities. Second, even the most likely mother-child genotype pair can have exclusion probabilities that are different from the average for the locus. Of course, all of the exclusion probabilities (minimum, most likely, maximum, or locus-specific average) are increasing functions of the heterozygosity at the locus. In these computations, mutations of the paternal allele are not considered. The same conclusions also hold when the true biological father is excluded because of mutations of the transmitted paternal alleles (see lower part of Table 1). Of course, the absolute paternity exclusion probabilities are all smaller by several orders of magnitude, if mutation is the only source of exclusion.

In Fig. 1 we show the dichotomy of the distributions of the number of loci with respect to which exclusions are to be found when the exclusion is truly due to non-paternity (blank histograms) versus when they are caused by mutation alone (shaded histograms). In panel (a) of Fig. 1, we plotted these distributions with locus-specific average exclusion probabilities ($P_E(l)$ and $\overline{P}_E(\mu)$), while in panel (b), the same computations are done for the most-likely values of $P_E(l)$ and $P_E(\mu)$ for each locus. For both panels, the dichotomy of the distributions is obvious; by non-paternity we expect exclusions at little more than 5 locus for this battery of 11 markers (mean = 5.28 with s.d. 1.61), while mutations alone will produce far less than one-locus exclusion (mean = 0.0004 with s.d. = 0.0192). In fact, with mutations alone, the probability of finding exclusions at two or more loci is less than 10^{-7}. This is true for the average locus-specific exclusion rates, as well as for the most likely genotypes for the mother-child pairs. The low mutation rates at the STR loci, furthermore, suggests that even one-locus exclusion is likely to be due to non-paternity, and may not be caused by mutation at the locus where exclusion is observed.

RFLP VNTR Loci

To examine whether mutation rates higher than the ones for the STR loci affect the above conclusions, we also applied the theory to VNTR loci where mutation rates are known to be more frequent than those at STR loci. Southern blot RFLP analysis of seven VNTR loci (D1S7, D2S44, D4S139, D5S110, D10S28, D14S13, and D17S79) constitute the most common battery of RFLP VNTR markers for parentage testing. Since discrete alleles at these loci cannot be identified by the Southern blot RFLP analysis, paternity exclusion probabilities, or paternity index computations at such loci are done by binned definition of alleles. Using the fixed bin allele frequencies of these loci in Caucasians ((22) and Dr. Bruce Budowle, personal communication), we computed the locus-specific average, minimum, maximum, and most-likely paternity exclusion probabilities that are shown in Table 2. Olaisen et al.

TABLE 1—*Locus-specific average, minimum, maximum and most-likely paternity exclusion probabilities for 11 STR loci caused by non-paternity (upper half) versus mutation of paternal alleles (lower half).*

Locus	Heterozygosity/ mutation rate	Probability of exclusion				Prob. M-C pair for most likely P_E
		Average	Minimum	Maximum	Most likely	
Due to non-paternity						
CD4	0.6786	0.3986	0.0938	0.9948	0.3752	0.2955
CSF1PO	0.7398	0.5030	0.1586	0.9942	0.4432	0.2599
CYAR04	0.7251	0.4890	0.0932	0.9895	0.4191	0.2721
F13A01	0.7308	0.4938	0.1092	0.9943	0.4293	0.2669
FABP	0.6471	0.4068	0.0713	0.9943	0.2278	0.3923
FESPFPS	0.8276	0.6587	0.2891	0.9960	0.5507	0.2086
LIPOL	0.6792	0.4167	0.0786	0.9529	0.3175	0.3291
PLA2A1	0.7236	0.5166	0.1553	0.9623	0.2924	0.3452
RENA4	0.3639	0.1901	0.0082	0.9944	0.0473	0.6493
TH01	0.7710	0.5600	0.1827	0.9893	0.4267	0.2682
TRINRAA	0.8223	0.6481	0.2865	0.9941	0.5407	0.2132
Due to mutation						
CD4	4.81×10^{-5}	3.29×10^{-5}	2.17×10^{-6}	3.05×10^{-3}	7.76×10^{-6}	0.2955
CSF1PO	4.88×10^{-5}	3.51×10^{-5}	3.68×10^{-6}	2.38×10^{-3}	9.24×10^{-6}	0.2599
CYAR04	4.64×10^{-5}	3.51×10^{-5}	1.56×10^{-6}	1.09×10^{-3}	6.89×10^{-6}	0.2721
F13A01	6.60×10^{-5}	5.17×10^{-5}	2.39×10^{-6}	2.54×10^{-3}	9.13×10^{-6}	0.2669
FABP	4.63×10^{-5}	3.40×10^{-5}	1.29×10^{-6}	2.32×10^{-3}	2.88×10^{-6}	0.3923
FESFPS	4.30×10^{-5}	3.57×10^{-5}	4.48×10^{-6}	1.80×10^{-3}	7.65×10^{-6}	0.2086
LIPOL	3.57×10^{-5}	2.21×10^{-5}	1.56×10^{-6}	2.86×10^{-4}	5.19×10^{-6}	0.3291
PLA2A1	4.61×10^{-5}	3.33×10^{-5}	3.38×10^{-6}	3.33×10^{-4}	4.19×10^{-6}	0.3452
RENA4	2.35×10^{-5}	1.47×10^{-5}	1.06×10^{-7}	2.07×10^{-3}	3.55×10^{-7}	0.6493
TH01	3.96×10^{-5}	2.83×10^{-5}	3.61×10^{-6}	1.04×10^{-3}	6.96×10^{-6}	0.2682
TRINRAA	6.13×10^{-5}	4.75×10^{-5}	8.40×10^{-6}	2.30×10^{-3}	1.39×10^{-5}	0.2132

NOTE: For the upper half, the second column represents the heterozygosity at the loci, whereas for the lower half, the mutation rates at the loci are listed (see text for source).

(23) and Eisenberg et al. (24) estimated mutation rates at these VNTR loci through direct observations on discordances of parental and children's genotype data. With these estimated mutation rates, we evaluated the rates of excluding a true father (average, minimum, maximum, and most likely) when exclusion is caused by mutations of paternal alleles at these loci. As in the case of the STR loci, here also we observe that the exclusionary chance due to mutations (combined $P_E(\mu)$ for 7 loci for a random mother child pair is 0.0580) is far less likely than those expected due to non-paternity (combined $P_E(C)$ for 7 loci for a random mother child pair is 0.999,996). Data presented in Table 2 also shows that even when the mutation rate at a locus is very high (e.g., D1S7), paternity exclusion is more likely due to non-paternity in comparison to that expected due to mutation. Of course, the higher mutation rates at these VNTR loci make the absolute probabilities of exclusion by mutation more frequent than that expected for STR loci.

In Fig. 2 we plot the distributions of the number of loci with respect to which an accused male will be excluded if he is truly not the father (blank histograms), versus when the true biological father is excluded because of mutations of the transmitting alleles for these 7 RFLP VNTR loci (shaded histograms). Panel (a) represents the computations for locus-averages, while in panel (b) the most-likely mother-child genotype pair is considered for the computations. The dichotomy of the distributions (exclusion due to nonpaternity versus exclusion due to mutations) is still obvious. For example, when all 7 loci are scored, the average number of loci showing paternity exclusions due to non-paternity is 5.69 (with s.d. = 1.00), while if exclusions are caused by mutation alone, on an average only 0.06 loci should show exclusion (s.d. = 0.24). The higher mutation rates at the VNTR loci make the single-locus exclusions comparatively more likely (than the STR loci) by mutations alone. For example, with these 7 loci, a true father will be excluded based on one of these seven loci (due to mutation) with an appreciable probability (0.058), while the chance of a similar event for the 11 STR loci is 0.0004 for a random mother-child pair. The appreciable probability of single-locus exclusion due to mutation, mainly arise from the hypermutability of the D1S7 locus. Nevertheless, the chance of exclusions at 2 or more loci is 0.0005 due to mutations, even when the hypermutable D1S7 locus is included in the analysis.

From both sets data (STR and VNTR) we observe that when hypervariable single locus probes are used for exclusion of paternity, exclusions at multiple loci cannot be explained by the high mutation rates at such loci. For VNTR loci, occasional single-locus exclusions may be caused by mutations. Irrespective of the use of specific mother-child genotype data, or the average exclusionary power of the loci, the suggestion of excluding a male from paternity by at least two-locus exclusions will avoid the possibility of excluding a true father. This suggestion is obviously a very conservative approach for the STR loci, since even one-locus exclusion by mutation at such loci can occur rarely in a population ($P < 0.0004$). Thus, we provide a rigorous statistical validation of the AABB prescription that exclusion of paternity should be based on when two or more loci show evidence of excluding the accused male. The above calculations also justify the avoidance of the hypermutable D1S7 locus for parentage analysis (which is generally the practice in US laboratories).

Discussions and Conclusions

The numerical illustrations, shown above, are for the allele frequency data from the US Caucasian populations. Since allele frequencies at STR and VNTR loci may be different for other populations, one should consider the general applicability of our

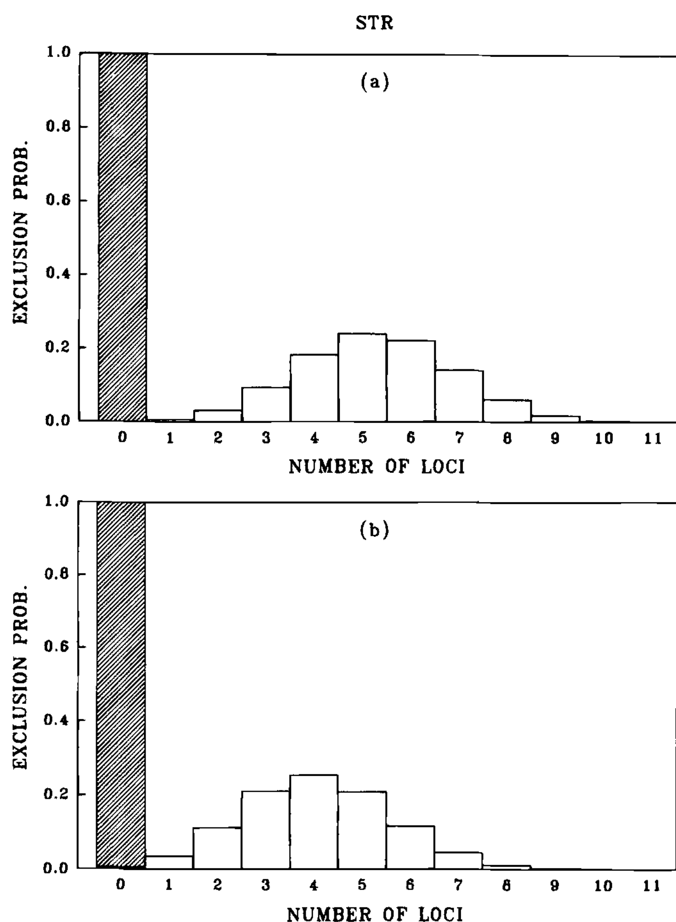

FIG. 1—*The distribution of the number of loci showing exclusions due to nonpaternity (blank bars) and due to mutations of parental alleles (shaded bars) for 11 autosomal short tandem repeat loci (see text for a list of loci). Allele frequency data for the loci are from US Caucasian populations (19,20).*

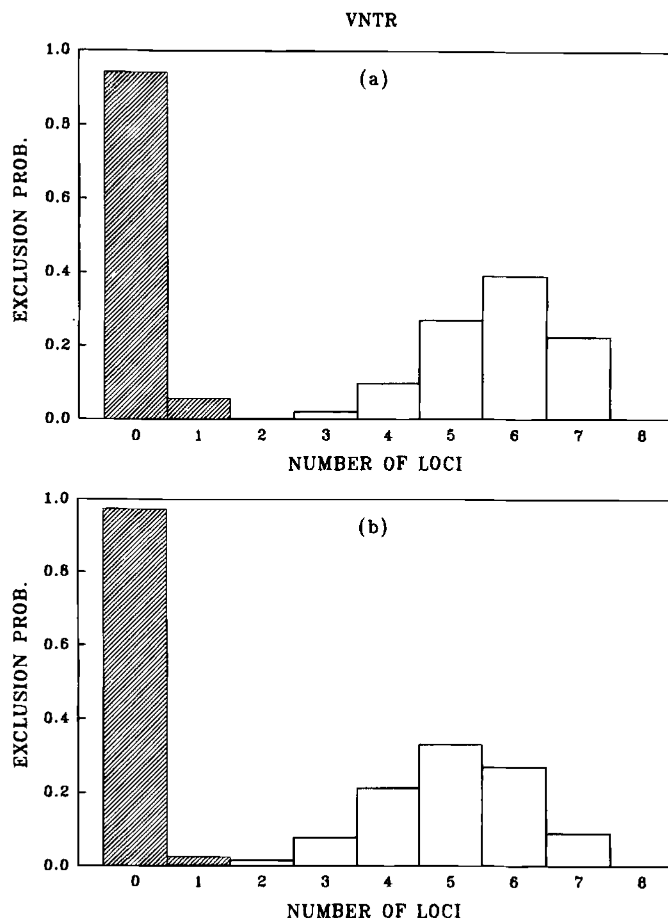

FIG. 2—*The distribution of the number of loci showing exclusions due to nonpaternity (blank bars) and due to mutations of parental alleles (shaded bars) for 7 VNTR loci (see text for a list of loci). Allele frequency data for the loci are from US Caucasian populations (22).*

TABLE 2—*Locus-specific average, minimum, maximum and most-likely paternity exclusion probabilities for 7 RFLP VNTR loci caused by non-paternity (upper half) versus mutation of paternal alleles (lower half).*

Locus	Heterozygosity/ mutation rate	Probability of exclusion				Prob. M-C pair for most likely P_E
		Average	Minimum	Maximum	Most likely	
Due to nonpaternity						
D1S7	0.9454	0.8893	0.7123	0.9920	0.8482	0.0733
D2S44	0.9261	0.8503	0.5914	0.9940	0.7674	0.1105
D4S139	0.8990	0.7978	0.4597	0.9920	0.6545	0.1615
D5S110	0.9301	0.8588	0.5944	0.9900	0.7797	0.1049
D10S28	0.9430	0.8842	0.6889	0.9880	0.8336	0.0801
D14S13	0.8986	0.8026	0.3944	0.9920	0.5960	0.1879
D17S79	0.7995	0.6044	0.2632	0.9940	0.5432	0.2120
(Due to mutation)						
D1S7	5.20×10^{-2}	1.98×10^{-2}	5.37×10^{-1}	4.76×10^{-2}	2.33×10^{-2}	0.0733
D2S44	3.40×10^{-4}	8.29×10^{-5}	5.36×10^{-3}	3.07×10^{-4}	1.00×10^{-4}	0.1105
D4S139	1.60×10^{-3}	2.69×10^{-4}	2.33×10^{-2}	1.41×10^{-3}	3.23×10^{-4}	0.1615
D5S110	8.00×10^{-3}	1.89×10^{-3}	7.20×10^{-2}	7.26×10^{-3}	2.42×10^{-3}	0.1049
D10S28	5.50×10^{-4}	1.94×10^{-4}	3.94×10^{-3}	5.01×10^{-4}	2.29×10^{-4}	0.0801
D14S13	7.20×10^{-4}	6.36×10^{-5}	7.44×10^{-3}	6.59×10^{-4}	7.84×10^{-5}	0.1879
D17S79	4.80×10^{-4}	3.99×10^{-5}	1.22×10^{-2}	4.06×10^{-4}	7.63×10^{-5}	0.2120

NOTE: For the upper half, the second column represents the heterozygosity for binned alleles, whereas for the lower half, the mutation rates at the loci are listed (see text for source).

results to parentage testing involving mothers and alleged fathers of differing racial origin. Analysis of allele frequency data at these loci from worldwide populations (25,26) show that the effect of inter-population allele frequency differences can be examined by choosing the most disparate population databases. The results of such an analysis are given in Table 3, where the basic summary statistics (i.e., exclusion probabilities based on at least one-locus and at least two-loci exclusions, caused by non-paternity as well as mutations), are compared for allele frequencies taken from the US Caucasian and US Black populations ((22) and Dr. Bruce Budowle, personal communication for the VNTR loci; and (19, 20) for the STR loci). Comparison of these statistics for these two databases indicate that the conclusion regarding the reliability of two or more exclusions by non-paternity is valid for the Black databases as well. Computations shown in this table also addresses another important issue raised by Pena (9) about the adequacy of single-locus probes alone for parentage analysis. Table 3 data shows that even with the stringency of deciding non- paternity based on two or more loci exclusions, for a battery of 18 VNTR and STR probes, the combined exclusion probability is 0.9999999, which is sufficiently large enough to resolve almost all paternity dispute cases. This probability is hardly affected by allele frequency differences between populations. Even when the hypermutable locus D1S7 is deleted, the combined probability of exclusion based on two or more locus exclusions does not reduce appreciably (0.9999996). Thus, while the multi-locus probes provide unique visual reliability of parentage determinations, combination of several single-locus probes provides almost as much accuracy as that of multi-locus probes. Use of single-locus probes has another distinct advantage, at least in societies where paternity dispute cases are litigated in a customarily adversarial jurisdiction system. The underlying assumptions, namely independence of alleles within and across loci, are readily testable for databases with single-locus probes, while such assumptions are somewhat difficult to test for multilocus probe data. Furthermore, when isolated loci show exclusions, the statistical consideration of whether such exclusions could arise from mutations of true paternal alleles can be based on the paternity index; the logic of this is discussed in the literature (27).

In the above theory we used the locus-specific exclusion probabilities assuming that the mother's genotypes are also known. Occasionally, paternity testing laboratories face situations where the mother is not available for testing (called deficiency cases). The above theory still holds; the only change required is in the evaluation of locus-specific exclusion probabilities. The average power of exclusion at a multi-allelic codominant autosomal locus, when the mother is not tested, is given by Garber and Morris (28), although an algebraic reduction of their formula (see the last equation of these authors) gives the expression $1 - 4a_2 + 4a_3 - 3a_4 + 2a_2^2$, where, as in equation (1), a_r represents the sum of r-th power of all allele frequencies at a locus. If an expression for $P_E(C)$ is required for a specific genotype of the child (but the mother is untested), equation (2) can be used without any change, since the obligatory paternal alleles consist of the collection of all alleles in the child's genotype. Analogous changes can be made for loci involving dominance relationships between alleles, although such situations are uncommon for DNA markers. However, DNA markers may occasionally demonstrate non-detectable alleles in RFLP analysis (29–32) with single-locus minisatellite probes, and PCR analysis of microsatellite CA- repeat loci (33,34). These situations are parallel to each locus of the HLA system, for which Chakravarti and Li (12) have given the required formula for exclusion probability. Although they showed that the occurrence of "non-detectable" alleles reduces the exclusion probability, the reduction is not critical for DNA markers for two reasons. First, even at the STR loci the number of segregating alleles are large enough to make the reduction trivial. Second, when the silent alleles are disregarded and all detectable alleles are assumed codominant, the computed allele frequencies become conservative (i.e., larger than expected). These offset the bias caused by over-reporting the exclusion probabilities. For example, when there are 11 equally frequent alleles at a locus, one of which is non-detectable, Chakravarti and Li's (12) computed average exclusionary power (with at least one-locus showing exclusion) at the locus is 0.775, whereas if each of the codominant equi-frequent 11 alleles are detectable, the average exclusion probability at the locus becomes 0.813.

In summary, the analyzes presented in this work suggest that a battery of single-locus probe RFLP markers involving 4 to 6 VNTR loci, along with PCR analysis of the STR loci, proves sufficiently reliable for parentage analysis. The reliability of inferring non-parentage is further increased when the two-or-more locus exclusion is required to establish non-paternity, since a person who is truly not the father of a child will be rarely excluded on the basis

TABLE 3—*Effect of allele frequency differences on exclusion probabilities due to nonpaternity and mutations.*

Loci	Number of loci	Population	$P_E(C)$	$P_E^{\ddagger}(C)$	Mean	(s.d.)
Due to nonpaternity						
VNTR	7	US Caucasians	0.9999957	0.999838	5.687	(1.004)
		US Blacks	0.9999994	0.999968	6.023	(0.903)
STR	11	US Caucasians	0.999446	0.993162	5.281	(1.605)
		US Blacks	0.998957	0.988547	4.976	(1.605)
Combined	18	US Caucasians	$\approx$ 1.0	0.9999999	10.969	(1.893)
		US Blacks	$\approx$ 1.0	0.9999999	10.999	(1.841)
Due to mutation						
VNTR	7	US Caucasians	0.0579	0.000529	0.058	(0.237)
		US Blacks	0.0581	0.000538	0.059	(0.237)
STR	11	US Caucasians	0.000370	$< 10^{-7}$	0.0004	(0.0192)
		US Blacks	0.000368	$< 10^{-7}$	0.0004	(0.0192)
Combined	18	US Caucasians	0.058276	0.000551	0.059	(0.238)
		US Blacks	0.058414	0.000559	0.059	(0.238)

NOTE: $P_E(C)$ = probability of exclusion based on at least one-locus exclusions for a random mother-child pair; $P_E^{\ddagger}(C)$ = probability of exclusion based on two or more loci exclusions for random mother-child pairs; mean = average number of loci for which the male will be excluded; s.d. = the standard deviation of the number of loci with respect to which the male is excluded.

of only one of such a system of loci, and exclusions at two or more loci are very unlikely to be caused by mutations of paternally transmitted alleles. With the stringency of exclusions at two or more loci, the loss of exclusionary power for the battery of markers is not severe enough to reduce the power of DNA testing. Furthermore, not using the D1S7 locus for parentage analysis is supported by statistical evidence. Inter-population alleles frequency differences do not effect any of these conclusions.

Acknowledgments

This work was supported by grants GM 41399 and GM 45861 from the National Institutes of Health. We thank Dr. A. Eisenberg for comments and suggestions, and Dr. Bruce Budowle was kind enough to supply the VNTR databases used in this analysis.

References

(1) Gottschall JL, Endean DJ. An overview of DNA analysis in paternity testing: current practices. In: Proceedings from the Second International Symposium on Human Identification 1991. Madison (WI): Promega Corporation, 1991:91–102.
(2) Pena SDJ, Chakraborty R. Paternity testing in the DNA era. Trends in Genetics 1994;10:204–09.
(3) Smouse PE, Chakraborty R. The use of restriction fragment length polymorphisms in paternity analysis. Am J Hum Genet 1986;38: 918–39.
(4) Morris JW, Sanda AI, Glassberg J. Biostatistical evaluation of evidence from continuous allele frequency distribution deoxyribonucleic acid (DNA) probes in reference to disputed paternity and identity. J Forensic Sci 1989;34:1311–17.
(5) Alford RL, Hammond HA, Coto I, Caskey CT. Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am J Hum Genet 1994;55:190–95.
(6) Jeffreys AJ, Royle NJ, Wilson V, Wong Z. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 1988;332:278–81.
(7) Chakraborty R, Schull WJ. A note on the distribution of the number of exclusions to be expected in paternity testing. Am J Hum Genet 1976;28:615–18.
(8) Chakraborty R. The distribution of the number of heterozygous loci in an individual in natural populations. Genetics 1981;98:461–66.
(9) Pena SDJ. Pitfalls of paternity testing based solely on PCR typing of minisatellites and microsatellites. Am J Hum Genet 1995;56:1503–04.
(10) Maha GC, Mason JM, Stuhlmiller GM, Heine U. Reply to Pena. Am J Hum Genet 1995;56:1505–06.
(11) Selvin S. Probability of nonpaternity determined by multiple allele codominant systems. Am J Hum Genet 1980;32:276–78.
(12) Chakravarti A, Li CC. The effect of linkage on paternity calculations. In: Walker RH, editor. Inclusion Probabilities in Parentage Testing. Arlington (VA): Am Association of Blood Banks, 1983;411–22.
(13) Chakraborty R, Meagher T, Smouse PE. Parentage analysis with genetic markers in natural populations. I. the expected proportion of offspring with unambiguous paternity. Genetics 1988;118:527–36.
(14) Chakraborty R, Ferrell RE. Correlation of paternity index with probability of exclusion and efficiency criteria of genetic markers for paternity testing. Forensic Sci Int 1982;19:113–24.
(15) Boyd WC. Tables and nomograms for calculating chances of excluding paternity. Am J Hum Genet 1954;26:426–33.
(16) Chakraborty R, Shaw MW, Schull WJ. Exclusion of paternity: the current state of the art. Am J Hum Genet 1974;26:477–88.
(17) Chakraborty R. Paternity testing with genetic markers: Are Y-linked genes more efficient than autosomal ones? Am J Med Genet 1985;21:297–305.
(18) Rothman ED, Neel JV, Hoppe FM. Assigning a probability of paternity in apparent cases of mutation. Am J Hum Genet 1981;33:617–28.
(19) Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 1992;12:241–53.
(20) Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
(21) Weber JL, Wong C. Mutation of human short tandem repeats. Hum Molecular Genet 1993;8:1123–28.
(22) Budowle B, Monson KL, Anoe K, Baechtel FS, Bergman D, Buel E, et al. A preliminary report on binned general population data on six VNTR loci in Caucasians, Blacks and Hispanics from the United States. Crime Laboratory Digest 1991;18:9–26.
(23) Olaisen B, Bekkemoen M, Hoff-Olsen P, Gill P. Human VNTR mutation and sex. Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ, editors. In: DNA Fingerprinting: State of the Science. Basel: Birkhäuser Verlag, 1993:63–69.
(24) Eisenberg AJ, Clement M, Bever RA, Creacy SD, Gaskill ME, Carlson DP, et al. Further characterization of the VNTR probe LH1 (D5S110) and applications for DNA typing. In: Bär W, Fromeni A, Rossi U, editors. Advances on Forensic Haemogenetics. Heidelberg: Springer-Verlag, 1993;19–21.
(25) Laboratory Division of the Federal Bureau of Investigation. VNTR Population Data: A Worldwide Study, Vol. I–V. Washington DC: US Department of Justice, 1993.
(26) Chakraborty R, Jin L, Zhoug Y, Deka R. Intra- and inter- population variation at VNTR, short tandem repeat and polymarker loci and their implications in forensic and paternity analysis. In: Proceedings from the Fifth International Symposium on Human Identification 1994. Madison (WI): Promega Corporation, 1995:29–41.
(27) Chakraborty R, Ryman N. Use of odds of paternity computations in determining the reliability of single exclusion in paternity testing. Human Heredity 1981;31:363–69.
(28) Garber RA, Morris JW. General equations for the average power of exclusion for genetic systems of n codominant alleles in one-parent and no-parent cases of disputed parentage. In: Walker RH, editor. Inclusion Probabilities in Parentage Testing. Arlington (VA): American Association of Blood Banks, 1983:277–80.
(29) Budowle B, Giusti AM, Waye JS, Baechtel FS, Fourney RM, Adams AE, et al. Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. Am J Hum Genet 1991;48:841–55.
(30) Jeffreys AJ, Royle NJ, Patel I, Armour JAL, MacLeod A, Collick A, et al. Principles and recent advances in human DNA fingerprinting. In: Burke T, Jeffreys AJ, Dolf G, Wolff R, editors. DNA Fingerprinting: Approaches and Applications. Basel: Birkhäuser Verlag, 1991:1–19.
(31) Steinberger EM, Thompson LD, Hartmann JM. On the use of excess homozygosity for subpopulation detection. Am J Hum Genet 1993;52:1275–77.
(32) Chakraborty R, Zhong Y, Jin L, Budowle B. Nondetectability of restriction fragments and independence of DNA fragment sizes within and between loci in RFLP typing of DNA. Am J Hum Genet 1993;55:391–401.
(33) Koorey DJ, Bishop GA, McCaughan GW. Allele non-amplification: a source of confusion in linkage studies employing microsatellite polymorphisms. Hum Molecular Genet 1993;2:289–91.
(34) Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR. Incidence and origin of 'null' alleles in the $(AC)_n$ microsatellite markers. Am J Hum Genet 1993;52:922–27.

Address requests for reprints or addition information to
Ranajit Chakraborty, Ph.D.
Human Genetics Center
School of Public Health
The University of Texas-Houston Health Science Center
P.O. Box 20334
Houston, Texas 77225